ECE 509: Convex Optimization	**Rutgers: Spring 2024**

Homework #5

March 7, 2024

Name: Ravi Raghavan *Extension: No*

Problem 1. *Steepest descent method in l_∞-norm. Explain how to find a steepest descent direction in the l_∞-norm, and give a simple interpretation.*

Solution. $\Delta x_{sd} = ||\nabla f(x)||_* \Delta x_{nsd}$
$x_{nsd} = argmin\{\nabla f(x)^T x \mid ||x||_\infty \leq 1\}$
$||\nabla f(x)||_* = sup\{\nabla f(x)^T x \mid ||x||_\infty \leq 1\}$

We know that the dual norm of the l_∞-norm is just the l_1-norm. Hence, we can say that $||\nabla f(x)||_*$, under the l_∞-norm, is just $||\nabla f(x)||_1$

When analyzing the normalized steepest descent direction, it is clear to see that $x_{nsd} = -sign(\nabla f(x))$. To clarify, the *sign* function is applied component-wise to $\nabla f(x)$

Hence, the unnormalized steepest descent direction is just $\Delta x_{sd} = -||\nabla f(x)||_1 sign(\nabla f(x))$

Simple Interpretation: If the partial derivative of the function f with respect to x_i is positive, we will have to reduce x_i. This makes sense because it will be like reducing f as well. On the other hand, if the partial derivative of the function f with respect to x_i is negative, we will increase x_i. This makes sense because it will be like reducing f as well.

Problem 2. *The pure Newton method.* Newton's method with fixed step size $t = 1$ can diverge if the initial point is not close to x^*. In this problem we consider two examples. Plot f and f', and show the first few iterates.

(a) $f(x) = \log(e^x + e^{-x})$ has a unique minimizer $x^* = 0$. Run Newton's method with fixed step size $t = 1$, starting at $x^{(0)} = 1$ and at $x^{(0)} = 1.1$

Solution. $\nabla(f(x)) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$
$\nabla^2(f(x)) = \frac{(e^x+e^{-x})^2 - (e^x - e^{-x})^2}{(e^x+e^{-x})^2} = \frac{4}{(e^x+e^{-x})^2}$
$x^{(0)} = 1$
$\Delta x_{nt} = -\frac{\nabla(f(x))}{\nabla^2(f(x))}$

$\Delta x_{nt} = -\frac{\nabla(f(x^{(0)}))}{\nabla^2(f(x^{(0)}))} = -\frac{0.761594155956}{0.419974341614} = -1.81343020392$

$x^{(1)} = x^{(0)} + \Delta x_{nt} = 1 - 1.81343020392 = -0.813430203924 = -8.134 \cdot 10^{-1}$

$x^{(2)} = x^{(1)} + \Delta x_{nt} = -0.813430203924 + 1.22283252051 = 0.409402316586 = 4.094 \cdot 10^{-1}$

$x^{(3)} = x^{(2)} + \Delta x_{nt} = 0.409402316586 - 0.456707233043 = -0.047304916457 = -4.730 \cdot 10^{-2}$

$x^{(4)} = x^{(3)} + \Delta x_{nt} = -0.047304916457 + 0.0473755192607 = 0.0000706028 = 7.060 \cdot 10^{-5}$

k	$x^{(k)}$	$f(x^{(k)})$	$f(x^{(k)}) - p^*$
1	$-8.134 \cdot 10^{-1}$	0.992869009359	0.299721828799
2	$4.094 \cdot 10^{-1}$	0.774710796746	0.0815636161863
3	$-4.730 \cdot 10^{-2}$	0.694265641074	0.00111846051368
4	$7.060 \cdot 10^{-5}$	0.693147183052	$2.4923778597 \cdot 10^{-9}$
5	$-2.346 \cdot 10^{-13}$	0.69314718056	$5.4733995 \cdot 10^{-14}$

Table 1: Newton's Method: $x^{(0)} = 1$

$x^{(5)} = x^{(4)} + \Delta x_{nt} = 0.0000706028 - 0.0000706028039346 = -2.346 \cdot 10^{-13}$

Now, let's change $x^{(0)} = 1.1$

$x^{(1)} = x^{(0)} + \Delta x_{nt} = -1.12855258527$

$x^{(2)} = x^{(1)} + \Delta x_{nt} = 1.23413113304$

$x^{(3)} = x^{(2)} + \Delta x_{nt} = -1.69516597992$

$x^{(4)} = x^{(3)} + \Delta x_{nt} = 5.71536010038$

$x^{(5)} = x^{(4)} + \Delta x_{nt} = -23021.3564857$

k	$x^{(k)}$	$f(x^{(k)})$	$f(x^{(k)}) - p^*$
1	-1.12855258527	1.22808384311	0.534936662546
2	1.23413113304	1.31546405981	0.622316879246
3	-1.69516597992	1.72830814921	1.03516096865
4	5.71536010038	5.71537095711	5.02222377655
5	-23021.3564857	23021.3564857	23020.6633385

Table 2: Newton's Method: $x^{(0)} = 1$

(b) $f(x) = -\log(x) + x$ has a unique minimizer $x^* = 1$. Run Newton's method with fixed step size $t = 1$, starting at $x^{(0)} = 3$

Solution. $\nabla(f(x)) = \frac{-1}{x} + 1 = \frac{x-1}{x}$
$\nabla^2(f(x)) = \frac{1}{x^2}$
$\Delta x_{nt} = -\frac{\nabla(f(x))}{\nabla^2(f(x))} = -(x-1)x$

We know that $x^{(0)} = 3$

$x^{(1)} = x^{(0)} + \Delta x_{nt} = -3$

Since $f(x) = -\log(x) + x$, we can see that $dom f = \{x \in (0, \infty)\}$ When we apply Newton's Method, we are already outside the domain.

Problem 3. *Gradient and Newton methods for composition functions.* Suppose $\phi : \mathbb{R} \to \mathbb{R}$ is increasing and convex, and $f : \mathbb{R}^n \to \mathbb{R}$ is convex, so $g(x) = \phi(f(x))$ is convex. (We assume that f and g are twice differentiable.) The problems of minimizing f and minimizing g are clearly equivalent.
Compare the gradient method and Newton's method, applied to f and g. How are the search directions related? How are the methods related if an exact line search is used? Hint. Use the matrix inversion lemma

Solution. Analysis of Gradient Descent and Newton's Method

Gradient Descent:
$\nabla g(x) = \phi'(f(x))\nabla f(x)$. It is clear to see that the $\nabla g(x)$ is a positive multiple of $\nabla f(x)$. Hence, if we use exact line search, the iterates of gradient descent will be the same. However, if we use another method such as backtracking search, it may not be the same.

Newton's Method:
$\nabla^2 g(x) = \phi^{''}(f(x))\nabla f(x)\nabla f(x)^T + \phi^{'}(f(x))\nabla^2 f(x)$

Search Direction when Newton's Method is applied to f: $\Delta x_{nt} = -\nabla^2 f(x)^{-1}\nabla f(x)$

Search Direction when Newton's Method is applied to g: $\Delta x_{nt} = -\nabla^2 g(x)^{-1}\nabla g(x) = -(\phi^{''}(f(x))\nabla f(x)\nabla f(x)^T + \phi^{'}(f(x))\nabla^2 f(x))^{-1}(\phi'(f(x))\nabla f(x))$

According to the Matrix Inversion Lemma, we know that $(A+BC)^{-1} = A^{-1} - A^{-1}B(I+CA^{-1}B)^{-1}CA^{-1}$

We see that $\Delta x_{nt} = -(\phi^{''}(f(x))\nabla f(x)\nabla f(x)^T + \phi^{'}(f(x))\nabla^2 f(x))^{-1}(\phi'(f(x))\nabla f(x))$

$A = \phi^{'}(f(x))\nabla^2 f(x)$
$B = \phi^{''}(f(x))\nabla f(x)$
$C = \nabla f(x)^T$

Let's simplify $(\phi^{''}(f(x))\nabla f(x)\nabla f(x)^T + \phi^{'}(f(x))\nabla^2 f(x))^{-1}$

$A^{-1} - A^{-1}B(I+CA^{-1}B)^{-1}CA^{-1}$
$C = \frac{1}{\phi''(f(x))}B^T$

$A^{-1} - A^{-1}B(I + \frac{1}{\phi''(f(x))}B^TA^{-1}B)^{-1}\frac{1}{\phi''(f(x))}B^TA^{-1}$

We know that B is a column vector. We also know that $\phi^{'}(f(x))$ is positive and $\nabla^2 f(x)$ is positive definite. Hence, A is positive definite.

Furthermore, since $\phi(x)$ is a convex function, $\phi^{''}(f(x)) \geq 0$. This means that $\frac{1}{\phi''(f(x))}B^TA^{-1}B$ is a positive number. Subsequently, $(I + \frac{1}{\phi''(f(x))}B^TA^{-1}B)^{-1}$ is a positive number as well. Hence, we can represent this positive number as k_1. We can see that $k_1 < 1$

$A^{-1} - k_1A^{-1}B\frac{1}{\phi''(f(x))}B^TA^{-1}$
$A^{-1} - \frac{k_1}{\phi''(f(x))}A^{-1}BB^TA^{-1}$

$x_{nt} = -(A^{-1} - \frac{k_1}{\phi''(f(x))}A^{-1}BB^TA^{-1})(\phi'(f(x))\nabla f(x))$

$x_{nt} = -(A^{-1} - \frac{k_1}{\phi''(f(x))}A^{-1}BB^TA^{-1})(\frac{\phi^{'}(f(x))}{\phi''(f(x))}B)$

$x_{nt} = -(\frac{\phi^{'}(f(x))}{\phi''(f(x))}A^{-1}B - \frac{k_1\phi^{'}(f(x))}{(\phi''(f(x)))^2}A^{-1}BB^TA^{-1}B)$

Since A is positive definite, $B^TA^{-1}B = k_2$ where k_2 is a positive constant

$x_{nt} = -(\frac{\phi^{'}(f(x))}{\phi''(f(x))}A^{-1}B - \frac{k_1k_2\phi^{'}(f(x))}{(\phi''(f(x)))^2}A^{-1}B)$

$x_{nt} = -((\frac{\phi^{'}(f(x))}{\phi''(f(x))} - \frac{k_1k_2\phi^{'}(f(x))}{(\phi''(f(x)))^2})(A^{-1}B))$

$x_{nt} = -((\frac{\phi^{'}(f(x))}{\phi^{''}(f(x))} - \frac{k_1 k_2 \phi^{'}(f(x))}{(\phi^{''}(f(x)))^2})(\frac{\phi^{''}(f(x))}{\phi^{'}(f(x))}\nabla^2 f(x)^{-1}\nabla f(x)))$

$x_{nt} = -(\nabla^2 f(x)^{-1}\nabla f(x)(1 - \frac{k_1 k_2}{\phi^{''}(f(x))}))$

$x_{nt} = -\nabla^2 f(x)^{-1}\nabla f(x)(1 - \frac{k_1 k_2}{\phi^{''}(f(x))})$

Let $k_3 = \frac{k_2}{\phi^{''}(f(x))}$. We see that $k_1 = (1 + k_3)^{-1}$. Hence, $k_1 k_3 < 1$. This means that x_{nt}, when Newton's method is applied on g, is a positive multiple of x_{nt}, when Newton's method is applied on f.

This indicates to us that, when we perform exact line search, the iterates should be the same for both f and g.

However, this is NOT guaranteed when we do backtracking line search.